

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2001, 15:23:33 : Search time 58.44 Seconds  
(without alignments)  
215.709 Million cell updates/sec

Title: US-09-593-316-4

Sequence: 1 MNKGVILSLMLVSTIVV.....IKLVKMSQTKRYNVRNV 368

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score  | Query Match | Length | ID | Description       |
|------------|--------|-------------|--------|----|-------------------|
| 1          | 2008   | 100.0       | 368    | 1  | GATR_BOVIN        |
| 2          | 1717.5 | 85.5        | 371    | 1  | P14769 bos taurus |
| 3          | 1476.5 | 73.5        | 394    | 1  | GATR_MOUSE        |
| 4          | 655    | 32.6        | 354    | 1  | PGAT_MOUSE        |
| 5          | 115.5  | 5.8         | 505    | 1  | AGAR_HUMAN        |
| 6          | 102.5  | 5.1         | 559    | 1  | PGAT_BOVIN        |
| 7          | 101.5  | 5.1         | 559    | 1  | PGAT_HUMAN        |
| 8          | 97.5   | 4.9         | 446    | 1  | GLNA_METM         |
| 9          | 95     | 4.7         | 1057   | 1  | POL_STVAT         |
| 10         | 94.5   | 4.7         | 559    | 1  | PGAT_RAT          |
| 11         | 94.5   | 4.7         | 2410   | 1  | POL1_BAYM         |
| 12         | 93     | 4.6         | 408    | 1  | YZ33_METJA        |
| 13         | 92.5   | 4.6         | 1418   | 1  | CA12_HUMAN        |
| 14         | 90     | 4.5         | 522    | 1  | HEX1_ENTH         |
| 15         | 90     | 4.5         | 708    | 1  | TALA_SV40         |
| 16         | 89.5   | 4.5         | 478    | 1  | MANC_ECOLI        |
| 17         | 89.5   | 4.5         | 2054   | 1  | YCF2_PINT         |
| 18         | 89     | 4.4         | 688    | 1  | TALA_POVIC        |
| 19         | 89     | 4.4         | 1826   | 1  | SUTS_HUMAN        |
| 20         | 88.5   | 4.4         | 369    | 1  | CA12_MOUSE        |
| 21         | 88.5   | 4.4         | 1459   | 1  | CA12_MOUSE        |
| 22         | 88     | 4.4         | 567    | 1  | YF28_METJA        |
| 23         | 87     | 4.3         | 754    | 1  | YVY1_GAEEL        |
| 24         | 86     | 4.3         | 1056   | 1  | POL_BIV06         |
| 25         | 86     | 4.3         | 1056   | 1  | POL_BIV06         |
| 26         | 86     | 4.3         | 1056   | 1  | POL_BIV06         |
| 27         | 85.5   | 4.3         | 385    | 1  | YGLM_YEAST        |
| 28         | 85.5   | 4.3         | 446    | 1  | GLNA_METM         |
| 29         | 85     | 4.2         | 541    | 1  | STR_HELPY         |
| 30         | 84     | 4.2         | 454    | 1  | GLNA_METJA        |
| 31         | 84     | 4.2         | 1018   | 1  | YGLA_METJA        |
| 32         | 84     | 4.2         | 1826   | 1  | SUTS_RAHIT        |
| 33         | 83.5   | 4.2         | 332    | 1  | METE_AERPE        |

## ALIGNMENTS

| RESULT 1 | ID  | GATR_BOVIN | STANDARD: | PRT: | 368 AA. |
|----------|---|------------|-----------|------|---------|
| AC       | P14769  |            |           |      |         |
| DT       | 01-APR-1990 (Rel. 14, Created)  |            |           |      |         |
| DT       | 01-APR-1990 (Rel. 14, Last sequence update)   |            |           |      |         |
| DT       | 30-MAY-2000 (Rel. 39, Last annotation update)   |            |           |      |         |
| DE       | N-ACETYLTRANSFERASE (UDP-GALACTOSE: BETA-D-GALACTOSYL-1,4-N-ACETYL- (GALACTOSYLTRANSFERASE) (UDP-GALACTOSE: BETA-D-GALACTOSYL-1,4-N-ACETYL- D-D-GLUCOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE). |            |           |      |         |
| DE       | D-D-GLUCOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE).   |            |           |      |         |
| GN       | GTA1.   |            |           |      |         |
| OS       | Bos taurus (Bovine).  |            |           |      |         |
| OC       | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;   |            |           |      |         |
| OC       | Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;   |            |           |      |         |
| OC       | Bovidae; Bovinae; Bos.  |            |           |      |         |
| OX       | NCBI_TaxID:9913;  |            |           |      |         |
| RP       | SEQUENCE FROM N.A.  |            |           |      |         |
| RX       | MEDLINE:89340543; PubMed:2503516;   |            |           |      |         |
| RA       | Shaper D.H., Shaper J.H., van den Bijnden D.H., van Tuenen A.J.,  |            |           |      |         |
| RT       | *Bovine alpha 1-->3-galactosyltransferase: isolation and  |            |           |      |         |
| RT       | characterization of a cDNA clone. Identification of homologous  |            |           |      |         |
| RT       | sequences in human genomic DNA.*;   |            |           |      |         |
| RL       | J. Biol. Chem. 264:14290-14297(1989).   |            |           |      |         |
| CC       | - FUNCTION: TRANSFER OF GALACTOSE FROM UDP-GALACTOSE TO AN  |            |           |      |         |
| CC       | ACCEPTOR MOLECULE (R).  |            |           |      |         |
| CC       | - CATALYTIC ACTIVITY: UDP-GALACTOSE + BETA-D-GALACTOSYL-1,4-  |            |           |      |         |
| CC       | N-ACETYL-D-GLUCOSAMINYL-R = UDP + ALPHA-D-GALACTOSYL-1,3-   |            |           |      |         |
| CC       | BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.  |            |           |      |         |
| CC       | - COFACTOR: MANGANESE.  |            |           |      |         |
| CC       | - PATHWAY: GLYCOSYLATION.   |            |           |      |         |
| CC       | - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND  |            |           |      |         |
| CC       | FORM IN TRANS CISTERNAE OF GOLGI.   |            |           |      |         |
| CC       | - DISEASE: AUTOIMMUNE DISEASE (ANTIBODIES AGAINST ENZYMATIC   |            |           |      |         |
| CC       | PRODUCT).   |            |           |      |         |
| CC       | - SIMILARITY: STRONG TO BLOOD GROUP AB TRANSFERASE.   |            |           |      |         |
| CC       | - SIMILARITY: STRUCTURAL SIMILARITY WITH THE OTHER MAMMALIAN  |            |           |      |         |
| CC       | GLYCOSYLTRANSFERASES.   |            |           |      |         |
| CC       | -----   |            |           |      |         |
| CC       | This SWISS-PROT entry is copyright. It is produced through a collaboration  |            |           |      |         |
| CC       | between the Swiss Institute of Bioinformatics and the EMBL collaboration -  |            |           |      |         |
| CC       | the European Bioinformatics Institute. There are no restrictions on its   |            |           |      |         |
| CC       | use by non-profit institutions as long as its content is in no way  |            |           |      |         |
| CC       | modified and this statement is not removed. Usage by and for commercial   |            |           |      |         |
| CC       | entities requires a license agreement (See <a href="http://www.jsb-sib.ch/announce/">http://www.jsb-sib.ch/announce/</a>  |            |           |      |         |
| CC       | or send an email to <a href="mailto:license@lsb-sib.ch">license@lsb-sib.ch</a> ).   |            |           |      |         |
| DR       | EMBL: J04989; AAA30558.1; ..  |            |           |      |         |
| DR       | PIR: A4785; A4785.  |            |           |      |         |
| KM       | Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;  |            |           |      |         |
| KM       | Signal-anchor; Golgi stack; Manganese.  |            |           |      |         |
| FT       | DOMAIN 1 6  |            |           |      |         |
| FT       | SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).   |            |           |      |         |
| FT       | DOMAIN 7 22   |            |           |      |         |
| FT       | LUMENAL, CATALYTIC (POTENTIAL).   |            |           |      |         |
| FT       | CARBOHYD 293 293  |            |           |      |         |
| FT       | N-LINKED (GLCNAC: ...) (POTENTIAL).   |            |           |      |         |



CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CC NCBI\_TaxID=10090;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90046769; PubMed=2510162;  
 RA Larsen R.D., Rajan V.P., Ruff M.M., Kukowska-Latallo J.,  
 RA Cummings R.D., Lowe J.B.;  
 RT Isolation of a cDNA encoding a murine  
 RT UDPgalactose:beta-D-galactosyl-1,4-N-acetyl-D-glucosaminide  
 RT alpha-1,3-galactosyltransferase: expression cloning by gene  
 RT transfer.\*;  
 RT Proc. Natl. Acad. Sci. U.S.A. 86:8227-8231(1989).  
 CC [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92184813; PubMed=1544928;  
 RA Jozaissse D.H., Shaper N.L., Kim D., van den Eljnden D.H., Shaper J.H.;  
 RT \*Murine alpha 1,3-galactosyltransferase. A single gene locus  
 RT specifies four isoforms of the enzyme by alternative splicing.\*;  
 RT J. Biol. Chem. 267:5534-5541(1992).  
 CC -1- FUNCTION: TRANSFER OF GALACTOSE FROM UDP-GALACTOSE TO AN  
 CC ACCEPTOR MOLECULE (R).  
 CC -1- CATALYTIC ACTIVITY: UDP-GALACTOSE + BETA-D-GALACTOSYL-1,4-  
 CC N-ACETYL-D-GLUCOSAMINYL-R -> UDP + ALPHA-D-GALACTOSYL-1,3-  
 CC BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.  
 CC -1- COFACTOR: MANGANESE.  
 CC -1- PATHWAY: GLYCOSYLATION.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND  
 CC FORM IN TRANS CISTERNAE OF GOLGI.  
 CC -1- SIMILARITY: STRONG, TO BLOOD GROUP AB TRANSFERASE  
 CC -1- SIMILARITY: STRUCTURAL SIMILARITY WITH THE OTHER MAMMALIAN  
 CC GLYCOSYLTRANSFERASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M26925; AAA37657.1; -;  
 DR EMBL: M85153; AAA37711.1; -;  
 DR PIR: A34417; A34417.  
 DR MCD: M61:95704; Gytal.  
 KM Signal-anchor: Glycosyltransferase; Glycoprotein; Transmembrane;  
 KM Signal-anchor: Golgi stack; Alternative splicing; Manganese.  
 FT DOMAIN 1 41 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 42 60 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
 FT DOMAIN 61 394 LUMENAL, CATALYTIC (POTENTIAL).  
 FT CARBOHYD 83 83 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 319 319 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT VARSPLIC 1 35 MISSING (IN SHORTER ISOFORM).  
 FT VARSPLIC 62 62 R -> SPDSFLMITYHTK (IN SHORTER ISOFORM).  
 SV SEQUENCE 394 AA: 46475 MW: 7766831640DIBBF7 CRC64:  
 Query Match 73.5%; Score 1476.5; DB 1; Length 394;  
 Best Local Similarity 72.2%; Pred No. 1-2e-111;  
 Matches 268; Conservative 41; Mismatches 47; Indels 15; Gaps 3;  
 1 MNVGRKVLISMLVSTVIVFWFYHSPEGLFWINPSNPEVSGSSITOKGMLEPRWFNN 60  
 DB 36 MNVGRKVLISMLVSTVIVFWFYHSPEGLFWINPSNPEVSGSSITOKGMLEPRWFNN 83  
 QY 61 GYH--EEGDINEERFORNEDE-SKLKLSDMFNPRKRPVNVMTMKKAPVVEGYTNAV 117  
 DB 84 GYHGYOEDNEVRERKRGNDRIEEFQJMDWPNPNRRDVLTVTWKAPVVEGYTNAV 143  
 QY 118 LUNYAKOKITGLVFAFGRYIEHLEDFLSANKHFMVGHVPIFYIVDDVSHMPK 177  
 DB 144 LKRYATOKLITGLVFAFGRYIEHLEDFLSANKHFMVGHVPIFYIVDDVSHMPK 203

QY 178 LGRLSRKVFVKIKRERKMODISMRRKKTIGEHIVAHIOHVDPLFCMDVDVFCDFKQVE 237  
 DB 204 LNPJHSLOVFEIRSEKRWODISMRRKKTIGEHIVAHIOHVDPLFCMDVDVFCDFKQVE 263  
 QY 238 TLGSEVQLOLQAMWYKADPNDFYERKRESAAVIFPGCEDFYHANAIFGGTPTVLANITQE 297  
 DB 264 TLGGLVQLOLQAMWYKADPNDFYERKRESAAVIFPGCEDFYHANAIFGGTPTVLANITRE 323  
 QY 298 CFKGLDKKNDIEAOWHDESHLNKPYLNPRTKIIISPYCOWYHIGLPAHILKVNWSW 357  
 DB 324 CFKGLDKKNDIEAOWHDESHLNKPYLNPRTKIIISPYCOWYHIGLPAHILKVNWSW 383  
 QY 358 TKENYVANNV 368  
 DB 384 TKENYVANNV 394  
 RESULT 4  
 BGAT\_HUMAN STANDARD; PRT; 354 AA.  
 AC P16442;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE HISTO-BLOOD GROUP ABO SYSTEM TRANSFERASE (NAGAT) [INCLUDES:  
 DE GLUCOPROTEIN-FUCOSYL-GALACTOSIDE ALPHA-N  
 DE ACETYLGALACTOSAMINYLTRANSFERASE (EC 2.4.1.40) (FUCOSYLGLYCOPROTEIN  
 DE ALPHA-N-ACETYLGALACTOSAMINYLTRANSFERASE) (HISTO-BLOOD GROUP A  
 DE TRANSFERASE) (A TRANSFERASE); GLYCOPROTEIN-FUCOSYL-GALACTOSIDE ALPHA-  
 DE GALACTOSYLTRANSFERASE (EC 2.4.1.37) (FUCOSYLGLYCOPROTEIN 3-ALPHA-  
 DE GALACTOSYLTRANSFERASE) (HISTO-BLOOD GROUP B TRANSFERASE) (B  
 DE TRANSFERASE)].  
 GN ABO.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RN 11  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLIN=90110098; PubMed=2104828;  
 RA Yamamoto F.-I., Marken J., Tsuji T., White T., Clausen H.,  
 RA Hakomori S.-I.;  
 RT \*Cloning and characterization of DNA complementary to human UDP-  
 RT GalNAc-4-epimerase (alpha 1-3-galNAc transferase (histo-blood  
 RT group A transferase) mRNA.\*;  
 RL J. Biol. Chem. 265:1146-1151(1990).  
 RN 121  
 RP SEQUENCE FROM N.A.  
 RX MEDLIN=90238543; PubMed=2333095;  
 RA Yamamoto F.-I., Clausen H., White T., Marken J., Hakomori S.-I.;  
 RT \*Molecular genetic basis of the histo-blood group ABO system.\*;  
 RL Nature 345:229-233(1990).  
 RN 133  
 RP CHARACTERIZATION.  
 RX MEDLIN=91035461; PubMed=2121736;  
 RA Yamamoto F.-I., Hakomori S.-I.;  
 RT \*Sugar-nucleotide donor specificity of histo-blood group A and B  
 RT transferases is based on amino acid substitutions.\*;  
 RL J. Biol. Chem. 265:19257-19262(1990).  
 CC -1- FUNCTION: THIS PROTEIN IS THE BASIS OF THE ABO BLOOD GROUP SYSTEM.  
 CC THE HISTO-BLOOD GROUP ABO INVOLVES THREE CARBOHYDRATE ANTIGENS: A,  
 CC B, AND H. A, B, AND AB INDIVIDUALS EXPRESS A GLYCOSYLTRANSFERASE  
 CC ACTIVITY THAT CONVERTS THE H ANTIGEN TO THE A ANTIGEN (BY ADDITION  
 CC OF UDP-GALNAc) OR TO THE B ANTIGEN (BY ADDITION OF UDP-GAL),  
 CC WHEREAS O INDIVIDUALS LACK SUCH ACTIVITY.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GALACTOSAMINE + GLYCOPROTEIN  
 CC ALPHA-L-FUCOSYL-(1,2)-D-GALACTOSE = UDP + N-ACETYL-ALPHA-D-  
 CC GALACTOSAMINYL-(1,3)-[ALPHA-L-FUCOSYL-(1,2)]-D-GALACTOSE.  
 CC -1- CATALYTIC ACTIVITY: UDP-GALACTOSE + GLYCOPROTEIN ALPHA-L-FUCOSYL-  
 CC (1,2)-D-GALACTOSE = UDP + GLYCOPROTEIN ALPHA-D-GALACTOSYL-(1,3)-  
 CC [ALPHA-L-FUCOSYL-(1,2)]-D-GALACTOSE.  
 CC -1- PATHWAY: GLYCOSYLATION.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND

FORM IN TRANS ESTERASE OF GOLI. SOLUBLE FORM IN BODY FLUIDS.  
 1. THE SOLUBLE FORM DERIVES FROM THE MEMBRANE FORM BY  
 1. PROTEOLYTIC PROCESSING.  
 1. POLYMERIZATION. THE SPOONER SHOWN IS THAT OF THE A TRANSFERASE. THE  
 1. FORM DIFFERS BY A FEW RESIDUES SUBSTITUTIONS, THE O PHENOTYPE IS  
 1. RESULT OF A SINGLE BASE FRAMESHIFT DETECTION IN THE N-TERMINAL  
 1. EXTREMITY OF THE GENE.  
 1. SIMILARITY: STRONG. TO N-ACETYLACTOSAMINIDE ALPHA-1,3-  
 1. GALACTOSYLTRANSFERASE (EC 2.4.1.151).  
 1. SIMILARITY: STRUCTURAL SIMILARITY WITH THE OTHER MAMMALIAN  
 1. GLYCOSYLTRANSFERASES.  
 1. This Swiss Prot entry is copyright. It is produced through a collaboration  
 1. between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 1. the European Bioinformatics Institute. There are no restrictions on its  
 1. use by non-profit institutions as long as its content is in no way  
 1. modified and this statement is not removed. Usage by and for commercial  
 1. entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 1. or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 1. EMBL: J05175; AAA6792.1;  
 1. PIR: A34933; A34933.  
 1. PIR: S09593; S09593.  
 1. MIM: 110300;  
 1. TRANSFOS: glycosyltransferase; Glycoprotein; Transmembrane;  
 1. Signal anchor; Golgi stack; Polymorphism; blood group antigen.  
 1. CHAIN 54 454  
 1. TRANSMEM 1 32  
 1. SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 1. (POTENTIAL).  
 1. LUMENAL, CATALYTIC (POTENTIAL).  
 1. N-LINKED (GLYCANC. ...) (POTENTIAL).  
 1. CH -> W.  
 1. /FTID-VAR\_003408.  
 1. P -> L (IN ALLELE A2).  
 1. /FTID-VAR\_003409.  
 1. R -> G (IN GROUP B TRANSFERASE).  
 1. /FTID-VAR\_003410.  
 1. G -> S (IN GROUP B TRANSFERASE).  
 1. /FTID-VAR\_003411.  
 1. L -> M (IN GROUP B TRANSFERASE).  
 1. IMPORTANT FOR THE SPECIFICITY).  
 1. /FTID-VAR\_003412.  
 1. G -> A (IN GROUP B TRANSFERASE).  
 1. IMPORTANT FOR THE SPECIFICITY).  
 1. /FTID-VAR\_003413.  
 1. R -> W (IN ALLELE B3 OF GROUP B  
 1. TRANSFERASE).  
 1. /FTID-VAR\_003414.  
 1. SEQUENCE 454 AA: 409-44 MW: 403016.6301608 CRC64:  
 1. Query Match 42.6% Score 655; DB 1; Length 354;  
 1. Best Local Similarity 45.3%; Pred. No. 1.4e-45;  
 1. Matches 124; Conservative 49; Mismatches 99; Indels 2; Gaps 2;  
 1. 94 REPVNMTKAWKAVVSTYNNKAVLNNYAKOKITVGLTFVAVGVYIEHYDEEFLTSANK 153  
 1. 1  
 1. 81 KRIVAVVTVMAVAVTWEIGFNTDILNROKRIQNTTIGLTVPAIKKVA-FLKLEFLTAKK 139  
 1. 1  
 1. 154 IRKAVHPIVPIYIKVAVSVKPLLELAPLRSKPVFKIKPKERMODISMMKRTGHEIVAH 213  
 1. 1  
 1. 140 HNVVHVRVIVYVPTQPAVVPVTLATGKQLSVLEKRAVKRWQDVSMRREMIISDCERR 199  
 1. 1  
 1. 214 IOLVHVFETFMVVDVQVQKFGVETLGRSAVDLOAMMYKADINDQETTERKKSAAVTPRG 273  
 1. 1  
 1. 200 FLSVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVV 259  
 1. 1  
 1. 274 ESHVYVHAAVIAETPTQVNTQTFKGIKDKKNDIEQMDHDESHLNKFLNKLTKIL 333  
 1. 1  
 1. 260 ESHVYVHAAVIAETPTQVNTQTFKGIKDKKNDIEQMDHDESHLNKFLNKLTKIL 319

334 SPECWDYH-IQLADIKIKVMSQTEYVAVRN 366  
 1  
 320 SPFLMDQGLGMVAVLKRLEFAVPRKNIQAVRN 353  
 1  
 1. RESULT 5  
 1. ACAR-ALFAT STANDARD; PRT; 505 AA.  
 1. AC P13734;  
 1. DT 01-JAN-1990 (Ref. 13, created)  
 1. DT 01-JAN-1990 (Ref. 13, last sequence update)  
 1. DT 15-JUL-1999 (Ref. 38, last annotation update)  
 1. DE ACAR-AGARASE PRECURSOR (EC 3.2.1.81).  
 1. CN ACRA.  
 1. OS Alteromonas atlantica (Pseudomonas atlantica).  
 1. OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;  
 1. OC Pseudomonas.  
 1. OX NCHI\_TaxID=288;  
 1. RN 11  
 1. RP SEQUENCE FROM N.A.  
 1. RX MEDLINE-89123082; PubMed-2914859;  
 1. RA Belas R;  
 1. RT "Sequence analysis of the acra gene encoding beta-agarase from  
 1. Pseudomonas atlantica".  
 1. RL J. Bacteriol. 171:602-605(1989).  
 1. CC 1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,3-BETA-D-GALACTOSIDIC LINKAGES  
 1. CC IN AGAROSE, GIVING THE TETRAMER AS THE PREDOMINANT PRODUCT.  
 1. CC 1- SUBCELLULAR LOCATION: SECRETED.  
 1. CC 1- NAME: NAME: worthington biochem manual;  
 1. WWW="http://www.worthington-biochem.com/manual//Acra.htm".  
 1. This Swiss Prot entry is copyright. It is produced through a collaboration  
 1. between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 1. the European Bioinformatics Institute. There are no restrictions on its  
 1. use by non-profit institutions as long as its content is in no way  
 1. modified and this statement is not removed. Usage by and for commercial  
 1. entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 1. or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 1. EMBL: M22725; AAA25696.1;  
 1. PIR: A32261; A32261.  
 1. DR Hydrolyase; Glycosidase; Signal.  
 1. FT SIGNAL 1 23  
 1. FT CHAIN 24 505  
 1. FT ACT\_SITE 200 200 BETA-ACARASE.  
 1. FT ACT\_SITE 422 422 PROTON DONOR (POTENTIAL).  
 1. FT ACT\_SITE 422 422 NUCLEOPHILE (POTENTIAL).  
 1. SEQUENCE 505 AA: 57617 MW: 668750.4862801 CRC64:  
 1. Query Match 5.8% Score 115.5; DB 1; Length 505;  
 1. Best Local Similarity 18.1%; Pred. No. 0.057;  
 1. Matches 78; Conservative 64; Mismatches 147; Indels 151; Gaps 18;  
 1. 6 KYHLSMVAIVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVV 65  
 1. 1  
 1. 59 DWQDEHDLKATKMERHLDVYFGDNAGTVMNFGALPDVANIYAVDPQNIKACQADQRTN 118  
 1. 1  
 1. 103 WKAVVVECTINKAVLAVNTYAKOKITVG-----LTVAVGVYI 140  
 1. 1  
 1. 119 W-----GOKSKALHJYDGRGLMIGGQPRVAVIANTSPQCGSSAWQAGDAVGF 170  
 1. 1  
 1. 141 EHYLEFELISANKHNVGH--PVIFYIM-----VQVSR-----MPYLE----- 178  
 1. 1  
 1. 171 GQVYNEFRSGADPVTKGLAVVYFEVLENYQVYTAHNLGLKQPIVPIIDITFINNV 230  
 1. 1  
 1. 179 -----GLPSVVFVKIKPKERMODISMMKRTGHEIVAHIOHEVDFRCM 224  
 1. 1  
 1. 231 ADAPKQHNTHIKIGFTVAFPIFEDKREFAKMKELKFLIDISGSHNVYVTHFPD-----L 286

QY 225 DVDQVQ-----DKFVETLGEV-----AQLQAMV--K 252  
 DB 287 EDNRKFGSLKLEATLMDIDYSLALGETKPHVISEYGRGRNEMENAPASLQDMWELKT 346  
 QY 253 ADNDNFYERKREKA-----AYIPFGGDDFYHAAIFG-----GTPVQVLTJOCFGLIK 304  
 DB 347 ASPLMLQFISRPUSLSTIPFVPI-----KALMKTADGTFPYWRLIRQO-----K 392  
 QY 305 DKKNDIEAQM 314  
 DB 393 EAPNTEGEM 402

RESULT 6  
 PACT\_BOVIN  
 ID PACT\_BOVIN STANDARD: PRT: 559 AA.  
 AC 007537;  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 39, Last annotation update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE POLYPEPTIDE N-ACETYLGLACTOSAMINYLTRANSFERASE (EC 2.4.1.41) (PROTEIN-  
 DE UDP ACETYLGLACTOSAMINYLTRANSFERASE) (UDP-GALNAC:POLYPEPTIDE, N-  
 DE ACETYLGLACTOSAMINYLTRANSFERASE) (GALNAC-T1).  
 GN GALNT1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 41-74.  
 RC TISSUE=Intestine, and Colostrum;  
 RA MEDLINE=93286099; PubMed=7685345;  
 RA Homa F.L., Hollander T., Lehman D.J., Thomsen D.R., Elhammer A.P.;  
 RT "Isolation and expression of a cDNA clone encoding a bovine UDP-  
 RT GALNAC:polypeptide N-acetylglactosaminyltransferase.";  
 RL J. Biol. Chem. 268:12609-12616(1993).  
 RN [2]  
 RP SEQUENCE OF 41-559 FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Placenta, and Colostrum;  
 RX MEDLINE=93366815; PubMed=8360184;  
 RX Hagen F.K., van Muyckhuysen B., Tabak L.A.;  
 RA "Purification, cloning, and expression of a bovine UDP-GALNAC:  
 RT polypeptide N-acetylglactosaminyltransferase.";  
 RL J. Biol. Chem. 268:18960-18965(1993).  
 CC - FUNCTION: THIS PROTEIN CATALYZES THE INITIAL REACTION IN O-LINKED  
 CC OLIGOSACCHARIDE BIOSYNTHESIS. THE TRANSFER OF AN N-ACETYL-D-  
 CC GALACTOSAMINE RESIDUE TO A SERINE OR THREONINE RESIDUE ON THE  
 CC PROTEIN RECEPTOR. THE ENZYME MAY PREFERENTIALLY TRANSFER TO  
 CC THREONINE.  
 CC - CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GALACTOSAMINE + POLYPEPTIDE =  
 CC UDP + N-ACETYL-D-GALACTOSAMINYL-POLYPEPTIDE.  
 CC - COFACTOR: MANGANESE AND CALCIUM.  
 CC - PATHWAY: GLYCOSYLATION.  
 CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.  
 CC - TISSUE SPECIFICITY: COLOSTRUM CONTAINS A SOLUBLE FORM.  
 CC - SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.  
 CC - This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L07780; AAA30532.1; -;  
 DR EMBL: L17437; AAA68489.1; -;  
 DR InterPro: IPR000772; -;  
 DR InterPro: IPR001173; -;  
 DR Pfam: PF000535; Glycos\_transf\_2; 1.  
 DR Pfam: PF00652; Richn\_B\_lectin; 1.

KW Transferase: Glycosyltransferase; Transmembrane; Signal-anchor;  
 KM Golgi stack; Glycoprotein; Manganese; Calcium.  
 FT PROPEP 1 40 REMOVED IN SOLUBLE POLYPEPTIDE.  
 FT CHAIN 41 559 N-ACETYLGLACTOSAMINYLTRANSFERASE.  
 FT POLYPEPTIDE N-  
 FT ACETYLGLACTOSAMINYLTRANSFERASE, SOLUBLE  
 FT FORM.  
 FT DOMAIN 1 8 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 9 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
 FT DOMAIN 29 559 LUMENAL, CATALYTIC (POTENTIAL).  
 FT CARBOHYD 95 95 N-LINKED (GLCNAC, ...) (POTENTIAL).  
 FT CARBOHYD 117 117 POTENTIAL.  
 FT CARBOHYD 118 118 POTENTIAL.  
 FT CARBOHYD 119 119 POTENTIAL.  
 FT CARBOHYD 141 141 N-LINKED (GLCNAC, ...) (POTENTIAL).  
 FT CARBOHYD 288 288 POTENTIAL.  
 FT CARBOHYD 552 552 N-LINKED (GLCNAC, ...) (POTENTIAL).  
 FT SEQUENCE 559 AA; 64192 MW; E3E538C4DE569840 CMC64;

Query Match 5.1%; Score 102.5; DB 1; Length 559;  
 Best local Similarity 20.5%; Pred. No. 0.72;  
 Matches 80; Conservative 52; Mismatches 131; Indels 127; Gaps 22;

QY 63 HEEDGIDNEKREGRNEDESKLSDWFPFK--RPVYVMTKMAKAPVWEGTYNAVLUN 120  
 DB 55 HEGPGEKRPVVIPEKIDKMKEMKINQFNLMASEMIALNSLPVRLGCKTVYPDN 114  
 QY 121 -----YAKOKITVCLIVEAV--GRYIEHYLEFLTSANKHHPVGFVYVMDV 170  
 DB 115 LPTISVVIFFHNAMSTLRLTHSVINRSPRHLEI-----VLVIDA 157  
 QY 171 SRMPLELGLPLRF-----KVKIKPKRWODISMRRK-----T 205  
 DB 158 SERDFLK-RPLESVYKLVKVPVAVIMEQR--SGLRARLKGAANSQGYITFLDAHCECT 215  
 QY 206 IG--EHLVAVHIOHEVPLFCMDVDVFODKF---GVE-TLGSVAQLOAMVKKADPNDF 258  
 DB 216 VGLPEPLARIKHDKRTVCPITDVISDTFELMAGSDMTYGGFNNKLNFRVYVQRHM 275  
 QY 259 TYERRESAAVIFG-----EGDFY-----YHAA--IFGSGTPVOVNIQEC-- 298  
 DB 276 --DRKKGDRI-LPVRRPTMAAGLFSIDROYFOEIGTYDAGMDJMGGENLEISPRIMQCG 332  
 QY 299 -----FKG-----ILKOKKNDIEQWDESHLNKYFLNKP--TK 331  
 DB 333 TLEIVTCSHGVFRKRPATPYTFPGGTGQILNKNRNLAEVWDE--FKNEFYIISPGVTK 390  
 QY 332 ILSPYCWQV-----HICLPADIKLVKMSW 356  
 DB 391 V-----DTGDISRILGLRHLKLCRPFSS 413

RESULT 7  
 PACT\_HUMAN  
 ID PACT\_HUMAN STANDARD: PRT: 559 AA.  
 AC Q10472;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE POLYPEPTIDE N-ACETYLGLACTOSAMINYLTRANSFERASE (EC 2.4.1.41) (PROTEIN-  
 DE UDP ACETYLGLACTOSAMINYLTRANSFERASE) (UDP-GALNAC:POLYPEPTIDE, N-  
 DE ACETYLGLACTOSAMINYLTRANSFERASE) (GALNAC-T1).  
 GN GALNT1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Salivary gland;  
 RX MEDLINE=96115928; PubMed=8690719;  
 RX Meurer J.A., Naylor J.M., Baker C.A., Thomsen D.R., Homa F.L.,

KA Ellhammer A.P.: "cDNA cloning, expression, and chromosomal localization of a human  
 RT UDP-galactose-4-epimerase, N-acetylglucosaminyltransferase.";  
 RL J. Biochem. 118:568-574(1995).  
 KM SEQUENCE FROM N.A.  
 RX MEDLINE 96025800; PubMed 7592619;  
 RA White T., Bennett E.P., Takio K., Sorensen T., Bonding N.,  
 KA Clausen H.:  
 RT "Partial cloning and cDNA cloning of a human UDP-N-acetyl-alpha-D-  
 galactose-4-epimerase/N-acetylglucosaminyltransferase.";  
 RL J. Biol. Chem. 270:24156-24165(1995).  
 CC FUNCTION: THIS PROTEIN CATALYZES THE INITIAL REACTION IN O-LINKED  
 CC GLYCOSYLATION BIOSYNTHESIS, THE TRANSFER OF AN N-ACETYL-D-  
 CC GALACTOSAMINE RESIDUE TO A SERINE OR THREONINE RESIDUE ON THE  
 CC PROTEIN RECEPTOR.  
 CC CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GALACTOSAMINE + POLYPEPTIDE -  
 CC UDP + N-ACETYL-D-GALACTOSAMINE-POLYPEPTIDE.  
 CC COFACTOR: MANGANESE AND CALCIUM.  
 CC PATHWAY: GLYCOSYLATION.  
 CC SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN, GOLGI.  
 CC SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.  
 CC -----  
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC The European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR PMID: 941514; Acc50427.1;  
 DR EMBL: X85018; CAA59480.1;  
 DR MIM: 602273;  
 DR INTERPRO: IPRO00772;  
 DR InterPro: IPRO0173;  
 DR Pfam: PF00535; GlycosylTransf\_2; 1;  
 DR Trnam: PF00532; RctnB\_Lectin; 1;  
 KM Transfase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 KM Golgi stack; Glycoprotein; Manganese; Calcium;  
 KM PROPEP 1 40  
 CC REMOVED IN SOLUBLE POLYPEPTIDE  
 CC N-ACETYLGALACTOSAMINYLTRANSFERASE  
 CC (BY SIMILARITY).  
 CC POLYPEPTIDE N-  
 CC ACETYLGALACTOSAMINYLTRANSFERASE, SOLUBLE  
 CC FORM.  
 CC CYTOPLASMIC (POTENTIAL).  
 CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
 CC LUMENAL, CATALYTIC (POTENTIAL).  
 CC N-LINKED (GLUCNA...) (POTENTIAL).  
 CC POTENTIAL.  
 CC CARBOHYD 117 117  
 CC CARBOHYD 118 118  
 CC POTENTIAL.  
 CC CARBOHYD 119 119  
 CC POTENTIAL.  
 CC CARBOHYD 141 141  
 CC N-LINKED (GLUCNA...) (POTENTIAL).  
 CC POTENTIAL.  
 CC CARBOHYD 288 288  
 CC N-LINKED (GLUCNA...) (POTENTIAL).  
 CC CARBOHYD 541 541  
 CC N-LINKED (GLUCNA...) (POTENTIAL).  
 CC CARBOHYD 552 552  
 CC N-LINKED (GLUCNA...) (POTENTIAL).  
 KM SEQUENCE 559 AA: 64219 MW: 60681106821E5B CPG64;  
 SU  
 Query Match 5.18; Score 101.5; DB 1; Length 559;  
 Best local similarity 20.88; Pred. No. 0.87;  
 Matches 79; Conservative 53; Mismatches 131; Indels 127; Gaps 22;

DB 158 SERPILK KPLESYVKKIKAVVIVIMFOR SCILTRIKAAVSKGVITETLADHCEI 215  
 QY 206 IC-ENIVAHITQHEVDLEFMDVQVPODKF ---GYE TLGESSVADLQAMVYKADIND 258  
 DB 216 VGMLEPIILARIKHDRIVVCPILIVLSIDDFEYVAGSMDIMYGGFNMKILNFWVIVPQEM 275  
 QY 259 TYERKRESAAVLPFC-----RDEY-----YHAA--FGGIPVQVLTITQEC 298  
 DB 276 --DRKKDRT-LPAPTPTMAGLPSIDRDYFGLGYACMDIMKGMLETSPTIWOCCG 342  
 QY 299 -----FKG-----LKKKKNDIFACWHDSEHINKYPLTKKP--TK 331  
 DB 343 TLEIVTGVCHVGRKATPTPTFGTGLINKNNRIALFVMDP--KRFVYIISPTVK 390  
 QY 342 ILSPVCMYD-----HGLPADIKLVKMSW 356  
 DB 391 V-----DYGDLSRVGLRIKIQKPSW 413  
 RESULT 8  
 GLNA\_METMP STANDARD; PRT; 446 AA.  
 ID GLNA\_METMP  
 AC 050648;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE GLUTAMINE SYNTHETASE (EC 6.3.1.2) (GLUTAMATE--AMMONIA LIGASE) (GS).  
 GN GLNA.  
 OS Methanococcus maripaludis.  
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcus;  
 OC Methanococcus.  
 OX NCBI\_TaxID=39152;  
 OX RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN IL;  
 RX MEDLINE 99084955; PubMed 9864348;  
 KA Cohen Kupiec R., Marx C.J., Leish J.A.:  
 RT "Function and regulation of glna in the methanogenic archaeon  
 RT Methanococcus maripaludis.";  
 RL J. Bacteriol. 181:256-261(1999).  
 CC -I- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + NH(4) + ADP + GLUTAMINE +  
 CC ORTHOPHOSPHATE.  
 CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).  
 CC -I- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC The European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC FMBL: AF062391; AAD04645.1;  
 DR InterPro: IPRO01691;  
 DR Pfam: PF00120; gln-synt; 1;  
 DR PROSITE: PS00180; GLNA\_1; 1;  
 DR PROSITE: PS00181; GLNA\_ATP; 1;  
 KM SEQUENCE 446 AA: 49421 MW: 106584106651122A CPG64;  
 SU  
 Query Match 4.98; Score 97.5; DB 1; Length 446;  
 Best local similarity 22.38; Pred. No. 1.47;  
 Matches 61; Conservative 41; Mismatches 112; Indels 59; Gaps 14;

DB 129 GGEFTVGPPEFFILKEDACGSWVPADAG-----YFDLEPLDGGCDIRKIVFALENL 182  
 QY 207 GCHT-----VAHIOHEVDLHFCMDV---DOVFQDKFGEVL----- 239  
 DB 183 GHVASHHHEVAGEGHEVDKFPADAVKTAADSVVTRKTKITKILAAODCKATFMPKPEFGI 242  
 QY 240 -DESAVAQIQAMWYKADPNFTYERREKESAAVTPFGSDGYFAALFGGPT--QYLVNTIQ 296  
 DB 243 NCSGMCHCHOSTIMLNGESPSY-----DESAQY-QLSETGMSYVAGILDHAKSIVAVNTPTV 296  
 QY 297 ECFKCIILKDKKNDIEAOWHDESHLNKYPFLNKP 329  
 DB 297 NSYKRLVPGYEFAPVNIAM---ANSNRSAIVRYP 326

## RESULT 9

POL\_SIVAI STANDARD; PRT: 1057 AA.  
 AC 002836:  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE POL. POLYPEPTIDE (CONTAINS: PROTEASE (RETROPEPTIN) (EC 3.4.23.16);  
 REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4))  
 GN POL.  
 OS Simian immunodeficiency virus (isolate AGM / clone CRT-1) (SIV-AGM).  
 OC Viruses: Retroviridae; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=31684;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91220680; PubMed=2024476;  
 RA Fomsgaard A., Hirsch V.M., Allan J.S., Johnson P.R.;  
 RT "A highly divergent proviral DNA clone of SIV from a distinct species  
 of African green monkey."  
 RL Virology 182:397-402(1991).  
 CC -I- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
 DETERMINED.  
 CC -I- MISCELLANEOUS: THIS IS AN AFRICAN GREEN MONKEY ISOLATE.  
 CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
 KNOWN AS THE RETROPEPTIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DB EMBL: M66437; AAA91973.2; -  
 DB EMBL: M58410; CAB26041.1; -  
 DB HSSP: P03366; IHMV.  
 DR MEROPS: A02.003; -  
 DR InterPro: IPR000477; -  
 DR InterPro: IPR001037; -  
 DR InterPro: IPR001584; -  
 DR InterPro: IPR001969; -  
 DR InterPro: IPR002156; -  
 DR Pfam: PF00552; Integrase\_Zn; 1.  
 DR Pfam: PF00075; rnaaseh; 1.  
 DR Pfam: PF00665; rve; 1.  
 DR Pfam: PF00077; rvp; 1.  
 DR Pfam: PF00078; ryl; 1.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE: PS50175; ASP\_PROT\_RETROV; 1.  
 KM AIDS: Polyprotein: Hydroxylase: Aspartyl protease: Endonuclease;  
 Nuclease: Transferase: RNA-directed DNA polymerase.  
 FT ACT\_SITE 114 114 BY SIMILARITY.  
 SO SEQUENCE 1057 AA; 120566 MW; 0550E954011FAA2C CRC64;

Query Match 4.78; Score 95; DB 1; Length 1057;

Best Local Similarity 22.68; Pred. No. 6.5; Mismatches 111; Indels 92; Gaps 19;

Matches 72; Conservative 44; Mismatches 111; Indels 92; Gaps 19;

QY 69 INEEKQNEDESESKIKLSDMFPKRPVYMTMKAPVWECTYRNNAVLDNYAKKRT 128  
 DB 199 IEFKVKQLEKEDKPKLKW-PLSRKLEALTEICKQMEEGKLSIGDENP----- 250  
 QY 129 VGLVFAVAGYIE---HYLEEF--LTSANKHFM-----VGHVLFYIMVDVSMPLIEL 178  
 DB 251 -NTVFALKKKKIKTQWRMLVDFRRLNKATQDFPVQGLPAPA-----GLQKKKQITVIDI 305  
 QY 179 G-----PL-----KSKFKFK-----PEKMQ-----DISMKKMTIGBH 210  
 DB 306 GDAYVSIPLCKEPRKYAFTIPSVNNTGPIKYOINCLQCKSPITFQNTAANILREI 365  
 QY 211 VAH-----IOHEVDLEFCMDVDQVFD-----KQVETLESVAQIOW-W- 250  
 DB 366 KRHTPGLEIQVMDMLASDHDETRHQVDIVRKMLLEKGLTPKKVGRPEPMW 425  
 QY 251 YKADPNFTYERREKESAAVTPFGSDGYFAALFGGPTPIVQVNIQCHKGI----- 302  
 DB 426 YKLHFNKMTINKIE-----LPPLDSEFWTVNKLQKVVG-----VINMASQIYPGIKTKHTG 475  
 QY 303 -LKKKKNDI-EAOWHDES 318  
 DB 476 AMLRQKKMLLEPIWTFEA 494

## RESULT 10

PAGT\_RAT STANDARD; PRT: 569 AA.

AC Q10473:  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE POLYPEPTIDE N-ACETYLGLYCOSAMINYLTRANSFERASE (EC 2.4.1.41) (PROTEIN-  
 UDP ACETYLGLYCOSAMINYLTRANSFERASE) (UDP-GALNAc:POLYPEPTIDE, N-  
 ACETYLGLYCOSAMINYLTRANSFERASE) (GALNAc-T1).  
 GN GALNT1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Sublingual gland;  
 RX MEDLINE=96318029; PubMed=8748168;  
 RA Hagen F., Gregorie C.A., Tabak L.A.;  
 RT \*Cloning and sequence homology of a rat UDP-galNAc:polypeptide N-  
 acetylglucosaminyltransferase.\*  
 RL Glycoconj. J. 12:901-909(1995).  
 CC -I- FUNCTION: THIS PROTEIN CATALYZES THE INITIAL REACTION IN O-LINKED  
 OLIGOSACCHARIDE BIOSYNTHESIS, THE TRANSFER OF AN N-ACETYL-D-  
 GALACTOSAMINE RESIDUE TO A SERINE OR THREONINE RESIDUE ON THE  
 PROTEIN RECEPTOR.  
 CC -I- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GALACTOSAMINE + POLYPEPTIDE =  
 UDP + N-ACETYL-D-GALACTOSAMINYL-POLYPEPTIDE.  
 CC -I- COFACTOR: MANGANESE AND CALCIUM.  
 CC -I- PATHWAY: GLYCOSYLATION.  
 CC -I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN, Golgi.  
 CC -I- TISSUE SPECIFICITY: HEART, BRAIN, SPLEEN, LIVER, SKELETAL MUSCLE  
 AND KIDNEY.  
 CC -I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----







RL FASEB J. 5:2052-2060(1991).  
 RN [11]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE-97255959; PubMed 9101290;  
 RA Kivimäki H., Trump G., Prockop D.J.;  
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
 associated collagen (type IX), and network-forming collagen (type X)  
 cause a spectrum of diseases of bone, cartilage, and blood vessels."  
 RL Hum. Mutat. 9:400-415(1997).  
 RN [12]  
 RP VARIANT SER-1074.  
 RX MEDLINE-90046999; PubMed 2572591;  
 RA Vissint H., D'Alessio M., Lee H., Ramirez F., Godfrey M.,  
 RA Hollister D.W.;  
 RT "Glycine to serine substitution in the triple helical domain of pro-  
 alpha 1 (I) collagen results in a lethal perinatal form of short-  
 limbed dwarfism."  
 RL J. Biol. Chem. 264:18265-18267(1989).  
 RN [13]  
 RP VARIANT SER-1095--TYR-1440 DEL.  
 RX MEDLINE-89266907; PubMed 2544071;  
 RA Lee H., Vissint H., Ramirez F., Rogers D., Kimoon D.;  
 RT "Identification of the molecular defect in a family with  
 spandyloepiphyseal dysplasia."  
 RL Science 244:978-980(1989).  
 RN [14]  
 RP VARIANT OSTEOARTHRITIS CYS-650.  
 RX MEDLINE-90370826; PubMed 1975694;  
 RA Ala-Kosko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;  
 RT "Single base mutation in the type II procollagen gene (COL2A1) as a  
 cause of primary osteoarthritis associated with a mild  
 chondrodysplasia."  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).  
 RN [15]  
 RP VARIANT GLY-14 VAL-717.  
 RX MEDLINE-91291136; PubMed 2064612;  
 RA Bateman J.F., Hamman M., Chan D., Cole M.G.;  
 RT "Characterization of a type I collagen alpha 2(I) glycine-586 to  
 valine substitution in osteogenesis imperfecta type IV. Detection of  
 the mutation and prenatal diagnosis by a chemical cleavage method."  
 RL Biochem. J. 276:765-770(1991).  
 RN [16]  
 RP VARIANT OSTEOARTHRITIS CYS-650.  
 RX MEDLINE-91084471; PubMed 1985108;  
 RA Eyre D.R., Weis M.A., Moskowitz R.W.;  
 RT "Cartilage expression of a type II collagen mutation in an inherited  
 form of osteoarthritis associated with a mild chondrodysplasia."  
 RL J. Clin. Invest. 87:457-461(1991).  
 RN [17]  
 RP VARIANT HYPERCHONDROGENESIS GLU-984.  
 RX MEDLINE-94054748; PubMed 1429602;  
 RA Houdart R., Tillet G.E., Wiles M.A., Gruber H.E., Rimoin D.L.,  
 RA Cohen D.H., Eyre D.R.;  
 RT "An amino acid substitution (Gly853-->Glu) in the collagen alpha  
 1(I) chain produces hyperchondrogenesis."  
 RL J. Biol. Chem. 267:22522-22526(1992).  
 RN [18]  
 RP VARIANT HYPERCHONDROGENESIS SER-705.  
 RX MEDLINE-92262484; PubMed 1374906;  
 RA Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,  
 RA Ramirez F., Vitale E., Lee H.;  
 RT "Characterization of a type II collagen gene (COL2A1) mutation  
 identified in cultured chondrocytes from human hyperchondrogenesis."  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).  
 RN [19]  
 RP VARIANT WS-II ASP-198.  
 RX MEDLINE-93404428; PubMed 8417498;  
 RA Koerker J., Rivaranti P., Badila L., Kaerlaenen H.,  
 RA Kivikko K.L., Prockop D.J., Ala-Kosko L.;  
 RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate  
 for glycine alpha 1-67 and that causes cataracts and retinal  
 detachment: evidence for molecular heterogeneity in the Wagner  
 syndrome and the Stickler syndrome (arthro-ophthalmopathy)."

RL Am. J. Hum. Genet. 53:55-61(1993).  
 RN [20]  
 RP VARIANT SEMD CYS-840.  
 RA Tillet G.E., Weis M.A., Lachman R.S., Cohen D.H., Rimoin D.L.,  
 RA Eyre D.R.;  
 RT "A dominant mutation in the type II collagen gene (COL2A1) produces  
 spandyloepiphyseal dysplasia (SEMD), Strudwick type."  
 RL Am. J. Hum. Genet. 53:A209-A209(1993).  
 RN [21]  
 RP VARIANT OSTEOARTHRITIS CYS-650.  
 RX MEDLINE-93282819; PubMed 8507190;  
 RA Holderbaum D., Malmend G.J., Moskowitz R.W., Haqqi T.M.;  
 RT "Human cartilage from late stage familial osteoarthritis transcribes  
 type II collagen mRNA encoding a cysteine in position 519."  
 RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).  
 RN [22]  
 RP VARIANT SEMD ARG-285.  
 RX MEDLINE-93252400; PubMed 8486475;  
 RA Viikola M., Rivaranti P., Vuorio A.F., Kallila L., Ala-Kosko L.,  
 RA Peltonen L.;  
 RT "A mutation in the amino-terminal end of the triple helix of type II  
 collagen causing severe osteochondrodysplasia."  
 RL Genomics 16:282-285(1993).  
 RN [23]  
 RP VARIANT SERC CYS-206.  
 RX MEDLINE-94063862; PubMed 8244341;  
 RA Williams C.J., Considine E.L., Knowlton R.G., Reigato A., Neumann G.,  
 RA Harizan D., Buxton P., Jimenez S.A., Prockop D.J.;  
 RT "Spandyloepiphyseal dysplasia and precocious osteoarthritis in a  
 family with an Arg75-->Cys mutation in the procollagen type II gene  
 (COL2A1)."  
 RL Hum. Genet. 92:499-506(1993).  
 RN [24]  
 RP VARIANT SERC CYS-920.  
 RX MEDLINE-93315508; PubMed 8325895;  
 RA Chan D., Taylor T.K.F., Cole M.G.;  
 RT "Characterization of an arginine 789 to cysteine substitution in  
 alpha 1 (I) collagen chains of a patient with spandyloepiphyseal  
 dysplasia."  
 RL J. Biol. Chem. 268:15238-15245(1993).  
 RN [25]  
 RP VARIANT SERC SER-1128.  
 RX MEDLINE-93140139; PubMed 8423604;  
 RA Cole M.G., Ball R.K., Rogers J.G.;  
 RT "The clinical features of spandyloepiphyseal dysplasia congenita  
 resulting from the substitution of glycine 947 by serine in the alpha  
 1(I) chain of type II collagen."  
 RL J. Med. Genet. 30:27-35(1993).

Query Match 4.68; Score 92.5; DH 1; Length 1418;

Best Local Similarity 18.78; Pred. No. 15;

Matches 45; Conservative 25; Mismatches 50; Indels 121; Gaps 10;

QY 23 EYHSPESL-----FWINPSR-----NPEVSSST-- 48  
 DB 1198 ESTSPESKRNPAATCKDKLICHPEKMSQDWMIDPNQCTLDMKVCNMECTCYVP 1257  
 QY 49 -----OKGWLPR-----WF-----NNGYHEEDGDIENEKRONELESKLKSDWNP 91  
 DB 1258 NPANPKKNWSSKSKKHHWFGELINGGPHFSYGDNLAPNTANVQMTFLRL----- 1311  
 QY 92 FKRPVVNTKKAIVVWVGYYNNAVLNYYAKOKITVGLIVFAVGRVLEHYLEFTTSA 151  
 DB 1312 -----LSTEGSON-----ITVICKNSIA-----YLDFAACNL 1338  
 QY 152 NKEWV-----GHPVPLTVVD-----VSRMPLLEACP 180  
 DB 1339 KALLIGSNQVETIAFGNSRFTYALKKGCKIKHGKMKCTVIFYRSOKTSPLDIPAP 1398  
 QY 181 L 181  
 DB 1399 M 1399

RESULT 14  
 ID HEXL ENTHI STANDARD; PRT; 522 AA.  
 AC P49009;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE BETA-HEXOSAMINIDASE (EC 3.2.1.52) (N-ACETYL-BETA-GLUCOSAMINIDASE)  
 DE (BETA-GLUCAMINASE) (BETA-N-ACETYLHEXOSAMINIDASE).  
 GN HEX-A1 OR HEX-1.  
 OS Entamoeba histolytica.  
 OC Eukaryota; Entamoebidae; Entamoeba.  
 OX NCBI\_TaxID=5759;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HM-1:IMSS;  
 RX MEDLINE=96010595; PubMed=7581340;  
 RA Beanan M.J., Bailey G.B.;  
 RT \*The primary structure of an Entamoeba histolytica  
 RT beta-hexosaminidase A subunit.\*;  
 RL J. Eukaryot. Microbiol. 42:632-636(1995).  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL NON-REDUCING N-ACETYL-  
 CC D-HEXOSAMINE RESIDUES IN N-ACETYL-BETA-D-HEXOSAMINIDES.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 20 OF GLYCOSYL HYDROLASES.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: U09735; AAAB0165.1;  
 DR HSSP; P06865; IQBC.  
 DR InterPro: IPR001540;  
 DR Pfam: PF00728; GLYCO\_hydro.20; 1.  
 DR PRINTS: PR00738; GLYHYDRLASE20.  
 KM Hydrolase; Glycosidase.  
 SU SEQUENCE 522 AA; 59436 MW; B793C4B565A83AB CRC64;

Query Match 4.5%; Score 90; DB 1; Length 522;  
 Best Local Similarity 20.5%; Pred. No. 6.7;  
 Matches 55; Conservative 38; Mismatches 101; Indels 74; Gaps 14;

QY 17 VIVFWEYIHSPE---GSLFWINPSRNP-----EYGGSSIOK 50  
 DB 246 VVANCMDYIVSTKRGENVLSINPA-NPRTFPIIALMKELSDTGTGYVHAGDEV-- 302  
 QY 51 GWWLPWFNGYHEHEDDINEKQPNEDSKLSDMFNRPEVYVMTKKAVYME 110  
 DB 303 -WTSQMSKS---KEVSDIQKFKSKGLN-SLTELEGYNKTAQOEVI--HNKHHVYME 354  
 QY 111 GTNRAVLN-----YAKQKITVGL-TVFVAGRIEYMLE-----EFLTSANKH 154  
 DB 355 EYFKKNAGKNTIIQYWDRIQLQYVNSGYKAIFFSAGFYLDKQMLCNSYSTCVNTH 414  
 QY 155 FVNGHVIPEYIVWDVSRMPLIELGLRSFKVF-----KIKPKRMQDISMMMKIT 206  
 DB 415 SM-----WWTNNDMTDNDPDKLSSEKENVIGGEGCSMGESTDEQNFEDRVFYSAI 469  
 QY 207 GEH-----IVAHIQHEV--DPLFCMDV 226  
 DB 470 AERLMSKESVVDKESHEVRANVLRCLDV 497

RESULT 15  
 ID TALA\_SV40 STANDARD; PRT; 708 AA.  
 AC P03070;  
 DT 21-JUL-1986 (Rel. 01, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE LARGE T ANTIGEN.  
 OS Simian virus 40 (SV40).  
 OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Polyomavirus.  
 OX NCBI\_TaxID=10633;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=78159686; PubMed=205947;  
 RA Reddy V.B., Thimmappa R., Dhar R., Subramanian K.N., Zain R.S.,  
 RA Pan J., Ghosh P.K., Cejma M.L., Weissman S.M.;  
 RT \*The genome of simian virus 40.\*;  
 RL Science 200:494-502(1978).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=776;  
 RX MEDLINE=79156432; PubMed=205802;  
 RA Fiers W., Contreas R., Haegeman G., Rogiers R., van de Voorde A.,  
 RA van Heuverswyn H., van Herreweghe J., Volckaert G., Ysebaert M.;  
 RT \*Complete nucleotide sequence of SV40 DNA.\*;  
 RL Nature 273:113-120(1978).  
 RN [3]  
 RP NUCLEAR LOCALIZATION SIGNAL.  
 RX MEDLINE=85074458; PubMed=6096007;  
 RA Kalderson D., Roberts B.L., Richardson W.D., Smith A.P.;  
 RT \*A short amino acid sequence able to specify nuclear location.\*;  
 RL Cell 39:499-509(1984).  
 RN [4]  
 RP ZINC-FINGER DOMAIN.  
 RX MEDLINE=91056604; PubMed=2173794;  
 RA Hoess A., Moorel I.F., Fanning E., Arthur A.K.;  
 RT \*The finger domain of simian virus 40 large T antigen controls  
 RT DNA-binding specificity.\*;  
 RL J. Virol. 64:6291-6296(1990).  
 RN [5]  
 RP ZINC-FINGER DOMAIN.  
 RX MEDLINE=91237835; PubMed=1851875;  
 RA Loebner G., Stenger J.E., Ray S., Parsons R.E., Anderson M.E.,  
 RA Tegtmeyer P.;  
 RT \*The zinc finger region of simian virus 40 large T antigen is needed  
 RT for hexamer assembly and origin melting.\*;  
 RL J. Virol. 65:3167-3174(1991).  
 RN [6]  
 RP ASSOCIATION WITH HOST RB PROTEIN.  
 RX MEDLINE=88270506; PubMed=2839300;  
 RA Decaprio J.A., Ludlow J.W., Figue J., Shew J.-Y., Huang C.-M.,  
 RA Lee W.-H., Marsilio E., Paucha E., Livingston D.M.;  
 RT \*SV40 large tumor antigen forms a specific complex with the product of  
 RT the retinoblastoma susceptibility gene.\*;  
 RL Cell 54:275-283(1988).  
 RN [7]  
 RP IN VITRO PHOSPHORYLATION.  
 RX MEDLINE=88265851; PubMed=2838952;  
 RA Graesser F.A., Scheidmann K.H., Tuazon P.T., Traub J.A., Walter G.;  
 RT \*In vitro phosphorylation of SV40 large T antigen.\*;  
 RL Virology 165:13-22(1988).  
 RN [8]  
 RP SEQUENCE OF 102-118, AND PHOSPHORYLATION SITE THR-124 BY CDC2.  
 RX MEDLINE=90015163; PubMed=2552322;  
 RA McVey D., Britzuela L., Mohr I., Marshak D.R., Gluzman Y., Brach D.;  
 RT \*Phosphorylation of large tumour antigen by cdc2 stimulates SV40 DNA  
 RT replication.\*;  
 RL Nature 341:503-507(1989).  
 RN [9]  
 RP IN VITRO ACTIVATION OF DNA REPLICATION BY PP2A.  
 RX MEDLINE=90059993; PubMed=2551176;  
 RA Virshup D.M., Kaufman M.G., Kelly T.J.;  
 RT \*Activation of SV40 DNA replication in vitro by cellular protein  
 RT phosphatase 2A.\*;  
 RL EMBO J. 8:3891-3898(1989).  
 RN [10]  
 RP REGULATION OF REPLICATION BY PHOSPHORYLATION.  
 RX MEDLINE=90263097; PubMed=2160857;

RA gives C.  
 RT \*The replication functions of SV40 T antigen are regulated by phosphorylation.\*  
 RL Cell 61:745-758(1990).  
 RN (11)  
 RP DEPHOSPHORYLATION BY HOST PP2A.  
 RX MEDLINE-91172186; PubMed-1848668;  
 RA Scheidtmann K.H., Mumby M.C., Kandel R., Walter G.;  
 RT \*Dephosphorylation of simian virus 40 large-T antigen and p53 protein by protein phosphatase 2A: inhibition by small-t antigen.\*  
 RL Mol. Cell. Biol. 11:1996-2003(1991).  
 RN (12)  
 RP SITES OF DEPHOSPHORYLATION BY HOST PP2A.  
 RX MEDLINE-91162764; PubMed-1448420;  
 RA Scheidtmann K.H., Virshup D.M., Kelly T.J.;  
 RT \*Protein phosphatase 2A dephosphorylates simian virus 40 large T antigen specifically at residues involved in regulation of DNA-binding activity.\*  
 RL J. Virol. 65:2098-2101(1991).  
 RN (13)  
 RP ACTIVATION OF REPLICATION BY PHOSPHORYLATION ON THR-124.  
 RX MEDLINE-96211524; PubMed-8648725;  
 RA Mevius D., Wachter K., Dejmeyer P.;  
 RT \*Mechanisms of simian virus 40 T-antigen activation by phosphorylation of threonine 124.\*  
 RL J. Virol. 70:3887-3893(1996).  
 CC -1- FUNCTION: INITIATES DNA UNWINDING AND REPLICATION VIA ELABORATE INTERACTIONS WITH THE VIRAL ORIGIN OF REPLICATION. BINDS TWO ADJACENT SITES IN THE SV40 ORIGIN.  
 CC -1- SUBUNIT: FORMS HOMODIMERS IN THE PRESENCE OF ATP.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DOMAIN: THE ZINC FINGER REGION CONTRIBUTES TO PROTEIN-PROTEIN INTERACTIONS ESSENTIAL FOR THE ASSEMBLY OF STABLE T-ANTIGEN HEXAMERS AT THE ORIGIN OF REPLICATION AND HEXAMERS ARE NEEDED FOR SUBSEQUENT ALTERATIONS IN THE STRUCTURE OF ORIGIN DNA.  
 CC -1- PTM: PHOSPHORYLATED ON BOTH SERINE AND THREONINE RESIDUES. PHOSPHORYLATION ON SER-120 AND SER-123 INHIBITS VIRAL REPLICATION, WHILE THAT ON THREONINE-124 ENHANCES REPLICATION BY ACTIVATING THE DNA BINDING DOMAIN.  
 CC -1- PTM: DEPHOSPHORYLATED PREFERENTIALLY BY PP2A ON SERINE RESIDUES 12, 124, 677 AND PERHAPS 679. SMALL T ANTIGEN INHIBITS THE DEPHOSPHORYLATION BY THE AC FORM OF PP2A.  
 CC -1- MISCELLANEOUS: THE INITIAL 82 RESIDUES OF THE LARGE T AND SMALL T ANTIGENS FROM SV40 ARE CODED BY THE SAME NOCTURNE SEQUENCE.  
 CC THIS SWISS PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announc/> or send an email to [licensing@isb-sib.ch](mailto:licensing@isb-sib.ch)).  
 CC EMBL: V01480; CAA24672.1;  
 CC EMBL: J02400; AAB59924.1;  
 CC F1R: A04607; T0V974.  
 CC TRANSFAC: T00788;  
 CC CLYCS001498; P04070;  
 CC InterPro: IPR001623;  
 CC Pfam: PF00226; Pfam: 1;  
 CC PROSITE: PS00076; DNAS\_2; 1;  
 CC K01491;  
 CC K01492;  
 CC K01493;  
 CC K01494;  
 CC K01495;  
 CC K01496;  
 CC K01497;  
 CC K01498;  
 CC K01499;  
 CC K01500;  
 CC K01501;  
 CC K01502;  
 CC K01503;  
 CC K01504;  
 CC K01505;  
 CC K01506;  
 CC K01507;  
 CC K01508;  
 CC K01509;  
 CC K01510;  
 CC K01511;  
 CC K01512;  
 CC K01513;  
 CC K01514;  
 CC K01515;  
 CC K01516;  
 CC K01517;  
 CC K01518;  
 CC K01519;  
 CC K01520;  
 CC K01521;  
 CC K01522;  
 CC K01523;  
 CC K01524;  
 CC K01525;  
 CC K01526;  
 CC K01527;  
 CC K01528;  
 CC K01529;  
 CC K01530;  
 CC K01531;  
 CC K01532;  
 CC K01533;  
 CC K01534;  
 CC K01535;  
 CC K01536;  
 CC K01537;  
 CC K01538;  
 CC K01539;  
 CC K01540;  
 CC K01541;  
 CC K01542;  
 CC K01543;  
 CC K01544;  
 CC K01545;  
 CC K01546;  
 CC K01547;  
 CC K01548;  
 CC K01549;  
 CC K01550;  
 CC K01551;  
 CC K01552;  
 CC K01553;  
 CC K01554;  
 CC K01555;  
 CC K01556;  
 CC K01557;  
 CC K01558;  
 CC K01559;  
 CC K01560;  
 CC K01561;  
 CC K01562;  
 CC K01563;  
 CC K01564;  
 CC K01565;  
 CC K01566;  
 CC K01567;  
 CC K01568;  
 CC K01569;  
 CC K01570;  
 CC K01571;  
 CC K01572;  
 CC K01573;  
 CC K01574;  
 CC K01575;  
 CC K01576;  
 CC K01577;  
 CC K01578;  
 CC K01579;  
 CC K01580;  
 CC K01581;  
 CC K01582;  
 CC K01583;  
 CC K01584;  
 CC K01585;  
 CC K01586;  
 CC K01587;  
 CC K01588;  
 CC K01589;  
 CC K01590;  
 CC K01591;  
 CC K01592;  
 CC K01593;  
 CC K01594;  
 CC K01595;  
 CC K01596;  
 CC K01597;  
 CC K01598;  
 CC K01599;  
 CC K01600;  
 CC K01601;  
 CC K01602;  
 CC K01603;  
 CC K01604;  
 CC K01605;  
 CC K01606;  
 CC K01607;  
 CC K01608;  
 CC K01609;  
 CC K01610;  
 CC K01611;  
 CC K01612;  
 CC K01613;  
 CC K01614;  
 CC K01615;  
 CC K01616;  
 CC K01617;  
 CC K01618;  
 CC K01619;  
 CC K01620;  
 CC K01621;  
 CC K01622;  
 CC K01623;  
 CC K01624;  
 CC K01625;  
 CC K01626;  
 CC K01627;  
 CC K01628;  
 CC K01629;  
 CC K01630;  
 CC K01631;  
 CC K01632;  
 CC K01633;  
 CC K01634;  
 CC K01635;  
 CC K01636;  
 CC K01637;  
 CC K01638;  
 CC K01639;  
 CC K01640;  
 CC K01641;  
 CC K01642;  
 CC K01643;  
 CC K01644;  
 CC K01645;  
 CC K01646;  
 CC K01647;  
 CC K01648;  
 CC K01649;  
 CC K01650;  
 CC K01651;  
 CC K01652;  
 CC K01653;  
 CC K01654;  
 CC K01655;  
 CC K01656;  
 CC K01657;  
 CC K01658;  
 CC K01659;  
 CC K01660;  
 CC K01661;  
 CC K01662;  
 CC K01663;  
 CC K01664;  
 CC K01665;  
 CC K01666;  
 CC K01667;  
 CC K01668;  
 CC K01669;  
 CC K01670;  
 CC K01671;  
 CC K01672;  
 CC K01673;  
 CC K01674;  
 CC K01675;  
 CC K01676;  
 CC K01677;  
 CC K01678;  
 CC K01679;  
 CC K01680;  
 CC K01681;  
 CC K01682;  
 CC K01683;  
 CC K01684;  
 CC K01685;  
 CC K01686;  
 CC K01687;  
 CC K01688;  
 CC K01689;  
 CC K01690;  
 CC K01691;  
 CC K01692;  
 CC K01693;  
 CC K01694;  
 CC K01695;  
 CC K01696;  
 CC K01697;  
 CC K01698;  
 CC K01699;  
 CC K01700;  
 CC K01701;  
 CC K01702;  
 CC K01703;  
 CC K01704;  
 CC K01705;  
 CC K01706;  
 CC K01707;  
 CC K01708;  
 CC K01709;  
 CC K01710;  
 CC K01711;  
 CC K01712;  
 CC K01713;  
 CC K01714;  
 CC K01715;  
 CC K01716;  
 CC K01717;  
 CC K01718;  
 CC K01719;  
 CC K01720;  
 CC K01721;  
 CC K01722;  
 CC K01723;  
 CC K01724;  
 CC K01725;  
 CC K01726;  
 CC K01727;  
 CC K01728;  
 CC K01729;  
 CC K01730;  
 CC K01731;  
 CC K01732;  
 CC K01733;  
 CC K01734;  
 CC K01735;  
 CC K01736;  
 CC K01737;  
 CC K01738;  
 CC K01739;  
 CC K01740;  
 CC K01741;  
 CC K01742;  
 CC K01743;  
 CC K01744;  
 CC K01745;  
 CC K01746;  
 CC K01747;  
 CC K01748;  
 CC K01749;  
 CC K01750;  
 CC K01751;  
 CC K01752;  
 CC K01753;  
 CC K01754;  
 CC K01755;  
 CC K01756;  
 CC K01757;  
 CC K01758;  
 CC K01759;  
 CC K01760;  
 CC K01761;  
 CC K01762;  
 CC K01763;  
 CC K01764;  
 CC K01765;  
 CC K01766;  
 CC K01767;  
 CC K01768;  
 CC K01769;  
 CC K01770;  
 CC K01771;  
 CC K01772;  
 CC K01773;  
 CC K01774;  
 CC K01775;  
 CC K01776;  
 CC K01777;  
 CC K01778;  
 CC K01779;  
 CC K01780;  
 CC K01781;  
 CC K01782;  
 CC K01783;  
 CC K01784;  
 CC K01785;  
 CC K01786;  
 CC K01787;  
 CC K01788;  
 CC K01789;  
 CC K01790;  
 CC K01791;  
 CC K01792;  
 CC K01793;  
 CC K01794;  
 CC K01795;  
 CC K01796;  
 CC K01797;  
 CC K01798;  
 CC K01799;  
 CC K01800;  
 CC K01801;  
 CC K01802;  
 CC K01803;  
 CC K01804;  
 CC K01805;  
 CC K01806;  
 CC K01807;  
 CC K01808;  
 CC K01809;  
 CC K01810;  
 CC K01811;  
 CC K01812;  
 CC K01813;  
 CC K01814;  
 CC K01815;  
 CC K01816;  
 CC K01817;  
 CC K01818;  
 CC K01819;  
 CC K01820;  
 CC K01821;  
 CC K01822;  
 CC K01823;  
 CC K01824;  
 CC K01825;  
 CC K01826;  
 CC K01827;  
 CC K01828;  
 CC K01829;  
 CC K01830;  
 CC K01831;  
 CC K01832;  
 CC K01833;  
 CC K01834;  
 CC K01835;  
 CC K01836;  
 CC K01837;  
 CC K01838;  
 CC K01839;  
 CC K01840;  
 CC K01841;  
 CC K01842;  
 CC K01843;  
 CC K01844;  
 CC K01845;  
 CC K01846;  
 CC K01847;  
 CC K01848;  
 CC K01849;  
 CC K01850;  
 CC K01851;  
 CC K01852;  
 CC K01853;  
 CC K01854;  
 CC K01855;  
 CC K01856;  
 CC K01857;  
 CC K01858;  
 CC K01859;  
 CC K01860;  
 CC K01861;  
 CC K01862;  
 CC K01863;  
 CC K01864;  
 CC K01865;  
 CC K01866;  
 CC K01867;  
 CC K01868;  
 CC K01869;  
 CC K01870;  
 CC K01871;  
 CC K01872;  
 CC K01873;  
 CC K01874;  
 CC K01875;  
 CC K01876;  
 CC K01877;  
 CC K01878;  
 CC K01879;  
 CC K01880;  
 CC K01881;  
 CC K01882;  
 CC K01883;  
 CC K01884;  
 CC K01885;  
 CC K01886;  
 CC K01887;  
 CC K01888;  
 CC K01889;  
 CC K01890;  
 CC K01891;  
 CC K01892;  
 CC K01893;  
 CC K01894;  
 CC K01895;  
 CC K01896;  
 CC K01897;  
 CC K01898;  
 CC K01899;  
 CC K01900;  
 CC K01901;  
 CC K01902;  
 CC K01903;  
 CC K01904;  
 CC K01905;  
 CC K01906;  
 CC K01907;  
 CC K01908;  
 CC K01909;  
 CC K01910;  
 CC K01911;  
 CC K01912;  
 CC K01913;  
 CC K01914;  
 CC K01915;  
 CC K01916;  
 CC K01917;  
 CC K01918;  
 CC K01919;  
 CC K01920;  
 CC K01921;  
 CC K01922;  
 CC K01923;  
 CC K01924;  
 CC K01925;  
 CC K01926;  
 CC K01927;  
 CC K01928;  
 CC K01929;  
 CC K01930;  
 CC K01931;  
 CC K01932;  
 CC K01933;  
 CC K01934;  
 CC K01935;  
 CC K01936;  
 CC K01937;  
 CC K01938;  
 CC K01939;  
 CC K01940;  
 CC K01941;  
 CC K01942;  
 CC K01943;  
 CC K01944;  
 CC K01945;  
 CC K01946;  
 CC K01947;  
 CC K01948;  
 CC K01949;  
 CC K01950;  
 CC K01951;  
 CC K01952;  
 CC K01953;  
 CC K01954;  
 CC K01955;  
 CC K01956;  
 CC K01957;  
 CC K01958;  
 CC K01959;  
 CC K01960;  
 CC K01961;  
 CC K01962;  
 CC K01963;  
 CC K01964;  
 CC K01965;  
 CC K01966;  
 CC K01967;  
 CC K01968;  
 CC K01969;  
 CC K01970;  
 CC K01971;  
 CC K01972;  
 CC K01973;  
 CC K01974;  
 CC K01975;  
 CC K01976;  
 CC K01977;  
 CC K01978;  
 CC K01979;  
 CC K01980;  
 CC K01981;  
 CC K01982;  
 CC K01983;  
 CC K01984;  
 CC K01985;  
 CC K01986;  
 CC K01987;  
 CC K01988;  
 CC K01989;  
 CC K01990;  
 CC K01991;  
 CC K01992;  
 CC K01993;  
 CC K01994;  
 CC K01995;  
 CC K01996;  
 CC K01997;  
 CC K01998;  
 CC K01999;  
 CC K02000;  
 CC K02001;  
 CC K02002;  
 CC K02003;  
 CC K02004;  
 CC K02005;  
 CC K02006;  
 CC K02007;  
 CC K02008;  
 CC K02009;  
 CC K02010;  
 CC K02011;  
 CC K02012;  
 CC K02013;  
 CC K02014;  
 CC K02015;  
 CC K02016;  
 CC K02017;  
 CC K02018;  
 CC K02019;  
 CC K02020;  
 CC K02021;  
 CC K02022;  
 CC K02023;  
 CC K02024;  
 CC K02025;  
 CC K02026;  
 CC K02027;  
 CC K02028;  
 CC K02029;  
 CC K02030;  
 CC K02031;  
 CC K02032;  
 CC K02033;  
 CC K02034;  
 CC K02035;  
 CC K02036;  
 CC K02037;  
 CC K02038;  
 CC K02039;  
 CC K02040;  
 CC K02041;  
 CC K02042;  
 CC K02043;  
 CC K02044;  
 CC K02045;  
 CC K02046;  
 CC K02047;  
 CC K02048;  
 CC K02049;  
 CC K02050;  
 CC K02051;  
 CC K02052;  
 CC K02053;  
 CC K02054;  
 CC K02055;  
 CC K02056;  
 CC K02057;  
 CC K02058;  
 CC K02059;  
 CC K02060;  
 CC K02061;  
 CC K02062;  
 CC K02063;  
 CC K02064;  
 CC K02065;  
 CC K02066;  
 CC K02067;  
 CC K02068;  
 CC K02069;  
 CC K02070;  
 CC K02071;  
 CC K02072;  
 CC K02073;  
 CC K02074;  
 CC K02075;  
 CC K02076;  
 CC K02077;  
 CC K02078;  
 CC K02079;  
 CC K02080;  
 CC K02081;  
 CC K02082;  
 CC K02083;  
 CC K02084;  
 CC K02085;  
 CC K02086;  
 CC K02087;  
 CC K02088;  
 CC K02089;  
 CC K02090;  
 CC K02091;  
 CC K02092;  
 CC K02093;  
 CC K02094;  
 CC K02095;  
 CC K02096;  
 CC K02097;  
 CC K02098;  
 CC K02099;  
 CC K02100;  
 CC K02101;  
 CC K02102;  
 CC K02103;  
 CC K02104;  
 CC K02105;  
 CC K02106;  
 CC K02107;  
 CC K02108;  
 CC K02109;  
 CC K02110;  
 CC K02111;  
 CC K02112;  
 CC K02113;  
 CC K02114;  
 CC K02115;  
 CC K02116;  
 CC K02117;  
 CC K02118;  
 CC K02119;  
 CC K02120;  
 CC K02121;  
 CC K02122;  
 CC K02123;  
 CC K02124;  
 CC K02125;  
 CC K02126;  
 CC K02127;  
 CC K02128;  
 CC K02129;  
 CC K02130;  
 CC K02131;  
 CC K02132;  
 CC K02133;  
 CC K02134;  
 CC K02135;  
 CC K02136;  
 CC K02137;  
 CC K02138;  
 CC K02139;  
 CC K02140;  
 CC K02141;  
 CC K02142;  
 CC K02143;  
 CC K02144;  
 CC K02145;  
 CC K02146;  
 CC K02147;  
 CC K02148;  
 CC K02149;  
 CC K02150;  
 CC K02151;  
 CC K02152;  
 CC K02153;  
 CC K02154;  
 CC K02155;  
 CC K02156;  
 CC K02157;  
 CC K02158;  
 CC K02159;  
 CC K02160;  
 CC K02161;  
 CC K02162;  
 CC K02163;  
 CC K02164;  
 CC K02165;  
 CC K02166;  
 CC K02167;  
 CC K02168;  
 CC K02169;  
 CC K02170;  
 CC K02171;  
 CC K02172;  
 CC K02173;  
 CC K02174;  
 CC K02175;  
 CC K02176;  
 CC K02177;  
 CC K02178;  
 CC K02179;  
 CC K02180;  
 CC K02181;  
 CC K02182;  
 CC K02183;  
 CC K02184;  
 CC K02185;  
 CC K02186;  
 CC K02187;  
 CC K02188;  
 CC K02189;  
 CC K02190;  
 CC K02191;  
 CC K02192;  
 CC K02193;  
 CC K02194;  
 CC K02195;  
 CC K02196;  
 CC K02197;  
 CC K02198;  
 CC K02199;  
 CC K02200;  
 CC K02201;  
 CC K02202;  
 CC K02203;  
 CC K02204;  
 CC K02205;  
 CC K02206;  
 CC K02207;  
 CC K02208;  
 CC K02209;  
 CC K02210;  
 CC K02211;  
 CC K02212;  
 CC K02213;  
 CC K02214;  
 CC K02215;  
 CC K02216;  
 CC K02217;  
 CC K02218;  
 CC K02219;  
 CC K02220;  
 CC K02221;  
 CC K02222;  
 CC K02223;  
 CC K02224;  
 CC K02225;  
 CC K02226;  
 CC K02227;  
 CC K02228;  
 CC K02229;  
 CC K02230;  
 CC K02231;  
 CC K02232;  
 CC K02233;  
 CC K02234;  
 CC K02235;  
 CC K02236;  
 CC K02237;  
 CC K02238;  
 CC K02239;  
 CC K02240;  
 CC K02241;  
 CC K02242;  
 CC K02243;  
 CC K02244;  
 CC K02245;  
 CC K02246;  
 CC K02247;  
 CC K02248;  
 CC K02249;  
 CC K02250;  
 CC K02251;  
 CC K02252;  
 CC K02253;  
 CC K02254;  
 CC K02255;  
 CC K02256;  
 CC K02257;  
 CC K02258;  
 CC K02259;  
 CC K02260;  
 CC K02261;  
 CC K02262;  
 CC K02263;  
 CC K02264;  
 CC K02265;  
 CC K02266;  
 CC K02267;  
 CC K02268;  
 CC K02269;  
 CC K02270;  
 CC K02271;  
 CC K02272;  
 CC K02273;  
 CC K02274;  
 CC K02275;  
 CC K02276;  
 CC K02277;  
 CC K02278;  
 CC K02279;  
 CC K02280;  
 CC K02281;  
 CC K02282;  
 CC K02283;  
 CC K02284;  
 CC K02285;  
 CC K02286;  
 CC K02287;  
 CC K02288;  
 CC K02289;  
 CC K02290;  
 CC K02291;  
 CC K02292;  
 CC K02293;  
 CC K02294;  
 CC K02295;  
 CC K02296;  
 CC K02297;  
 CC K02298;  
 CC K02299;  
 CC K02300;  
 CC K02301;  
 CC K02302;  
 CC K02303;  
 CC K02304;  
 CC K02305;  
 CC K02306;  
 CC K02307;  
 CC K02308;  
 CC K02309;  
 CC K02310;  
 CC K02311;  
 CC K02312;  
 CC K02313;  
 CC K02314;  
 CC K02315;  
 CC K02316;  
 CC K02317;  
 CC K02318;  
 CC K02319;  
 CC K02320;  
 CC K02321;  
 CC K02322;  
 CC K02323;  
 CC K02324;  
 CC K02325;  
 CC K02326;  
 CC K02327;  
 CC K02328;  
 CC K02329;  
 CC K02330;  
 CC K02331;  
 CC K02332;  
 CC K02333;  
 CC K02334;  
 CC K02335;  
 CC K02336;  
 CC K02337;  
 CC K02338;  
 CC K02339;  
 CC K02340;  
 CC K02341;  
 CC K02342;  
 CC K02343;  
 CC K02344;  
 CC K02345;  
 CC K02346;  
 CC K02347;  
 CC K02348;  
 CC K02349;  
 CC K02350;  
 CC K02351;  
 CC K02352;  
 CC K02353;  
 CC K02354;  
 CC K02355;  
 CC K02356;  
 CC K02357;  
 CC K02358;  
 CC K02359;  
 CC K02360;  
 CC K02361;  
 CC K02362;  
 CC K02363;  
 CC K02364;  
 CC K02365;  
 CC K02366;  
 CC K02367;  
 CC K02368;  
 CC K02369;  
 CC K02370;  
 CC K02371;  
 CC K02372;  
 CC K02373;  
 CC K02374;  
 CC K02375;  
 CC K02376;  
 CC K02377;  
 CC K02378;  
 CC K02379;  
 CC K02380;  
 CC K02381;  
 CC K02382;  
 CC K02383;  
 CC K02384;  
 CC K02385;  
 CC K02386;  
 CC K02387;  
 CC K02388;  
 CC K02389;  
 CC K02390;  
 CC K02391;  
 CC K02392;  
 CC K02393;  
 CC K02394;  
 CC K02395;  
 CC K02396;  
 CC K02397;  
 CC K02398;  
 CC K02399;  
 CC K02400;  
 CC K02401;  
 CC K02402;  
 CC K02403;  
 CC K02404;  
 CC K02405;  
 CC K02406;  
 CC K02407;  
 CC K02408;  
 CC K0240